THE VALUE OF INTER-ANNUAL CORRELATIONS

If $x_1, x_2, x_3 \cdots x_n$ be measures taken on the n individuals of a series in a given year and $x_1', x_2', x_3' \cdots x_n'$ be similar measures taken in a subsequent year, the correlation between the first and second measures on the same individual $r_{xx'}$, may be designated as a direct inter-annual correlation.[1] The purpose of this review is to illustrate the usefulness of such constants, with a view to extending their application, by bringing together examples of inter-annual correlations from various fields.

The immediate value of such coefficients may be purely scientific, economic, or both theoretical and practical.

Practically such means of prediction as correlation and regression formulæ should find wide application in breeding operations where it is desirable to weed out or send to the butcher at the earliest possible moment those individuals which can not be kept with the maximum profit. If the correlation between the egg production of a fowl in her pullet year and her laying capacity in any subsequent year be high, it is clear that those which on the average are to prove unprofitable may be sent to the pot when most desirable for that purpose, and before they have consumed two or more years' feed without yielding the maximum return in eggs. If, on the contrary, there be no correlation, the labor of selection in the pullet year is an unnecessary expense. If a cow's milking capacity be closely correlated with her milking record in her heifer year, the culling of dairy herds may be profitably carried out in the first year. In plant breeding experiments, involving either sexual or vegetative reproduction, selection of individuals for future propagation must be made, and at as early a date as possible. If the future yield per plant of hay can be estimated with considerable accuracy from a first year's culture the process of selecting clonal strains can be carried out with far greater rapidity than if one must wait for the results of subsequent years' tests. In all such cases the finality of a first judgment must depend in large degree upon the closeness of correlation between the results of successive experiments—in short upon the value of the inter-annual correlation coefficient.

[1] Cross inter-annual correlations in which the measures taken are of a different sort are sometimes useful, but examples of such are not considered in this review.

In dealing with egg production Pearl and Surface[2] give as the correlation in number of eggs for first and second year

$$r = .032 \pm .083,$$

a value which, though positive, is clearly insignificant with regard to its probable error.

Thus in this particular case the performance of the first year furnishes no clue to that of the second. With respect to egg-laying capacity, the record of the pullet year furnishes no criterion for elimination from the flock.

For milk yield in cattle the case seems to be quite different. Gavin[3] has found that there is a medium correlation between (*a*) the "revised maximum"[4] yield in quarts of successive lactations, and (*b*) between the "revised maximum" of the individual lactation periods and the highest revised maximum reached by the animal.

So far as I am aware the only worker who has published correlations between the characters of the same plant individuals in different years is Clark[5] whose results have been noted in these pages by Pearl.[6]

The correlation tables and constants show that plants of a given class in any year (height or weight of hay produced) will be highly variable in a subsequent year, but will on the average deviate from the mean of the whole culture of the year in the same direction and to about half the extent of the type selected in the preceding year. Thus if selection were made on the basis of a single year's test only, many individual plants of low yield would be discarded which in a subsequent year would have taken higher rank, while high-yielding plants would be retained which subsequently would give disappointing results. On the whole, however, the yield of a hay plant one year does furnish a valuable index to its yield in a subsequent year.

[2] Pearl, R. and F. M. Surface, "A Biometrical Study of Egg Production in the Domestic Fowl," I, *Bull. Bu. Anim. Ind.*, 110, 66, 1909.

[3] Gavin, Wm., "Studies in Milk Records: On the Accuracy of Estimating a Cow's Milking Capacity by Her First Lactation Period." *Jour. Agr. Sci.*, *5*, 377–390, 1913.

[4] "Revised Maximum" milk yield is the maximum day yield which is three times reached or exceeded in a lactation.

[5] Clark, C. F., "Variation and Correlation in Timothy," *Bull. Cornell Agr. Exp. Sta.*, 279, 1910.

[6] Pearl, R., AMER. NAT., *45*, 418–419, 1911.

That we are dealing with a real measure of the relatively permanent differentiation of individuals, and not with merely temporary differences due to growth, is indicated by the fact that the correlations between a first and a third year are about the same as those between a first and a second or a second and a third.

In other fields of plant industry such methods may be profitably applied. For example Sievers[7] after discussing at some length the question of the differentiation of belladonna plants with respect to alkaloidal content, warns the reader that "the investigation has hardly progressed far enough to yield any definite conclusions" but says in summarizing his data:

> A considerable number of plants with leaves rich in alkaloids in one season are found to have equally rich leaves in the following season. Furthermore, they frequently manifest the same characteristics at the various stages of growth during the season in comparison with other plants. The same facts are true with regard to plants which bear leaves with a low percentage of alkaloids.

How much more definite is the information conveyed by the simple statement that the inter-annual correlation[8] between the alkaloidal content for 1911 and 1912 is

$$r = .513 \pm .066!$$

Such studies as those by Stockberger on individual performance in hops[9] may be facilitated by the use of inter-annual correlation coefficients. He gives only the extremes of his series of individuals, but from these the correlations between yield per hill for different years are:

	Lowest Hills	Highest Hills
1909 and 1910	$.29 \pm .17$	$.59 \pm .13$
1910 and 1911	$.55 \pm .13$	$.52 \pm .14$
1909 and 1911	$.43 \pm .15$	$.30 \pm .18$

Such constants, deduced from materials which almost certainly

[7] Sievers, A. F., "Individual Variation in the Alkaloidal Content of Belladonna Plants," *Jour. Agr. Res.*, *1*, 129–146, 1913.

[8] In computing this coefficient a number of inconsistencies in the data table were discovered. The constant as given is probably as nearly correct as can be found from the available data.

[9] Stockberger, W. W., "A Study of Individual Performance in Hops," *Prac. Amer. Breed. Ass.*, 7, 452–457, 1912.

do not show the full strength of the correlation, remove at once all question concerning the relatively permanent differences in productiveness of the individual hills.

Consider next an illustration from hybridization of measurable characters.

Goodspeed and Clauson[10] have given the mean values of measurements of the flowers of individual plants of *Nicotiana* hybrids cultivated in 1912 and of corollas of the same plants cut back and flowered in 1913. The correlations between the mean dimensions for the two years I find to be:

N. Tabacum var. *macrophylia* ♀ × *N. sylvestris* ♂

F_1 plants, $N = 21$.

For spread of corolla, $r = .044 \pm .147$.

For length of corolla, $r = .169 \pm .143$.

Hybrid produced by crossing F_1 of the hybrid *N. Tabacum* "Maryland" ♀ by *N. Tabacum* ♂, with *N. sylvestris*, $N = 19$.

For spread of corolla, $r = .560 \pm .106$.

For length of corolla, $r = .788 \pm .059$.

These correlations show at once the high degree of uniformity of the F_1 of the first as compared with that of the second series. In all four cases the signs of the coefficients are positive, but those of the first class are insignificant in comparison with their probable errors. In both cases length of corolla is more closely correlated than breadth. Possibly this is due to errors of sampling only, or to greater difficulty in obtaining an exact measure of the spread of the limb. It may, however, indicate that some characters are more sharply and permanently differentiated from individual to individual than others.

That the latter may sometimes be the case is clearly shown by unpublished data of my own for the ligneous perennials *Staphylea trifolia* and *Hibiscus Syriacus*.[11]

[10] Goodspeed, J. H., and R. E. Clauson, "Factors Influencing Flower Size in *Nicotiana* with Special Reference to Questions of Inheritance," *Amer. Jour. Bot.*, *2*, 232–274, 1915.

[11] The constants are based in all cases on mean values of the characters of ovaries of shrubs well established in the Missouri Botanical Garden. In such work the number of individuals can never for practical reasons be very large, if a fairly large number of countings be made for each shrub. Furthermore much of the work which one does may be lost by some accident which precludes the securing of countings from each individual every year. If an individual is not represented in both of a pair of years it must be omitted entirely.

The accompanying tables show the correlations deduced for the characters indicated.[12]

INTER-ANNUAL CORRELATIONS FOR FRUITS OF STAPHYLEA

Relationship	Correlation for 1906 and 1908, $n = 19$	Correlation for 1907 and 1908, $n = 15$	Correlation for 1908 and 1909, $n = 20$
Ovules and ovules	.445 ± .069	.816 ± .058	.872 ± .036
Seeds and seeds	.063 ± .154	.064 ± .173	.056 ± .150
Asymmetry and asymmetry	.748 ± .068	.102 ± .172	.205 ± .145
Locular composition and locular composition	.601 ± .099	.294 ± .159	.335 ± .134

INTER-ANNUAL CORRELATION FOR FRUITS OF HIBISCUS, $n = 23$.

Relationship	Correlation for 1907 and 1908
Sepals and sepals	.451 ± .112
Bracts and bracts	.836 ± .042
Ovules and ovules	.941 ± .016
Seeds and seeds	.630 ± .085
Asymmetry and asymmetry	.747 ± .062
Locular composition and locular composition	.725 ± .067
Fertility and fertility	.610 ± .088
Correlation and correlation	.035 ± .141

The constants are very irregular in magnitude, but are without exception positive in sign. In many instances they are large. Thus in these individual shrubs which taxonomically show no differences[13] there is nevertheless a distinct differentiation in respect of the great majority of the characters examined.

While the probable errors are large the evidence warrants the conclusion that some are decidedly more highly correlated than others.

[12] Sepals = mean number of sepals in calyx.

Bracts = mean number of bracts in involucre.

Ovules = mean number of ovules formed per fruit.

Seeds = mean number of seeds matured per fruit.

Asymmetry = average radial asymmetry in the distribution of the number of ovules per locule. For method of computation see *Biometrika*, Vol. VII, pp. 477–478, 1910, and AMER. NAT., Vol. XLVI, p. 480, 1912.

Locular composition = average number of locules per fruit with an odd number of ovules. See citations above.

Fertility = coefficient of fertility (mean seeds per fruit) (mean ovules per fruit).

Correlation = coefficient of correlation between number of ovules and number of seeds per locule.

[13] I believe one of the *Hibiscus* shrubs had lighter flowers than the rest.

In *Hibiscus* the differentiation of the individuals with respect to number of bracts seems to be greater than that for number of sepals. For both *Staphylea* and *Hibiscus* the correlation for ovules is generally high. It is in every instance higher than that for mean number of seeds matured per fruit. Correlation for both mean number of seeds per fruit and relative number of seeds matured has a moderately large value in *Hibiscus,* but in *Staphylea* it is sensibly 0. In both species such characteristics of the ovary as radial asymmetry and locular composition seem to be rather sharply differentiated from individual to individual. This is probably due in part to differentiation with respect of number of ovules per fruit, but further discussion of the problem would be out of place in a note, the only purpose of which is to call attention to the usefulness, in both applied and pure science, of a quantitative means of detecting and expressing permanent differentiation.

In this brief review I have made no attempt to discuss fully all the biological phases of the problems suggested. The analysis of the data may in several instances be carried much further by the use of the statistical tools. Perhaps enough has been said to indicate that inter-annual coefficients may be of real service in practical animal husbandry, in plant breeding and in morphology and physiology. More than usefulness is not to be expected of any method.

J. ARTHUR HARRIS

THE PHENOMENON OF SELF STERILITY

In my paper which appeared in THE AMERICAN NATURALIST, Vol. XLIX, p. 79, the last seven lines on page seventy-nine should read as follows:

> Self-sterile plants crossed with self-sterile plants gave only self-sterile offspring. Certain self-fertile plants, however, gave only self-fertile offspring either when self-pollinated or when crossed with self-sterile plants. Other self-fertile plants gave ratios of 3 self-fertile to 1 self-sterile offspring when self-pollinated, and ratios of 1:1 when crossed with pollen from self-sterile, etc.

E. M. EAST.